DOCUMENT RESUME

ED 303 506 TM 012 751

AUTHOR Kirby, Peggy C.
TITLE Reflective Teaching and Teacher Effectiveness: Measurement Considerations.
PUB DATE Nov 88
NOTE 15p.
AVAILABLE FROM Department of Educational Leadership and Foundations, University of New Orleans, New Orleans, LA 70148.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Elementary Secondary Education; Evaluation Methods; *Factor Analysis; *Measurement Techniques; Objective Tests; Problem Solving; Teacher Attitudes; *Teacher Effectiveness; Teaching Styles; *Test Construction; Testing; Test Reliability; Test Validity
IDENTIFIERS *Reflective Teaching Instrument

ABSTRACT

An attempt was made to develop an objective instrument to assess a teacher's perceived engagement in reflective practice. A Reflective Teaching Instrument (RTI) was developed around three dimensions of reflective practice in teaching: (1) diagnosis (problem setting); (2) testing; and (3) personal causation. Indicators of each of the dimensions were compiled from a review of the literature to generate an instrument. Four educators comprised an expert panel that assessed the face validity of each item. The pilot instrument of 48 Likert-format items was administered to 40 practicing teachers enrolled in graduate classes. A field test was subsequently conducted with 102 public elementary and junior high school teachers, representing a response rate of 94%. Items empirically supported through factor analysis after the field test were retained, resulting in a revised 15-item scale. Reliability coefficients of the revised scale indicated that items appeared to measure the identified constructs. Further research should include refinement of the instrument with attention to the diagnosis subscale, which was of somewhat suspect construct validity. Research to support the relationship between scores on this instrument and teacher effectiveness is also needed. Factor structure and alpha reliability data of the RTI are tabulated. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED303506

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

PEGGY C. KIRBY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Reflective Teaching and Teacher Effectiveness: Measurement Considerations

Peggy C. Kirby

Department of Educational Leadership and Foundations

University of New Orleans

New Orleans, Louisiana 70148

November 1988

Running head: REFLECTIVE TEACHING INSTRUMENT

TM012751

In his address to the American Educational Research Association in April 1987, Donald Schon advocated reflective practice in teacher education. One year earlier in his presidential address, David Berliner noted differences between expert and novice teachers in their approaches to practice. He claimed that expert teachers possess a tacit knowledge of classrooms derived from experience that influences the manner in which they organize their lessons and pace instruction. This ability of the professional to integrate experience with theory and research in the formulation of solutions to unique problems of practice that are complex or unique is what Schon termed reflective practice.

A theory of reflective practice serves educators well. First, it assumes that professional knowledge does exist in fields such as education that are plagued by ambiguous or conflicting theories; and, secondly, it presumes that application of this knowledge requires more than imitation of experts. In short, the theory of reflective practice refutes the model of technical rationality which assumes that effective practice is the result of application of accepted techniques to recurring problems of practice. Reflective practice supports the notion of teaching as a profession.

Donald Schon (1983) proposes the theory of "reflection-in-action" that distinguishes the expert from the novice in situations of practice. Teachers do not learn to teach solely by imitating experts or by guided remediation of past errors, but are themselves researchers "continually reframing their world of work in response to puzzling or surprising events of practice" (Russell, 1985, p. 16).

Schon's (1983) theory of practice seeks to explain effectiveness in professionals confronted with a public crisis in confidence and with the professional pluralism that emerges in response to uncertainty and complexity. Reflective practice has become a popular catch phrase in education, but definition of the term and recognition of the use of reflective practice has eluded the educational researcher.

A handful of studies have emerged that examine the use of reflective practice by teachers (e.g., Korthagen, 1985; Russell, 1986; MacKinnon, 1986; Oberg & Field, 1986). All use the case study method and qualitative data analysis. Although there is evidence of a relationship between reflective practice and teacher effectiveness, this relationship has not been satisfactorily demonstrated partly

because no objective instrument exists to detect reflective practice.

This study is an initial attempt to develop an objective instrument that assesses a teacher's perceived engagement in reflective practice. Such an instrument would allow the theory of reflective practice, i.e., the relationship between reflective and effective practice, to be empirically tested.

THEORETICAL FRAMEWORK

The theoretical framework and indicators of reflective practice used in development of a Reflective Teaching Instrument (RTI) were derived from Argyris and Schon's (1974) earlier model of effective practice. This model assumes that technical rationality which views intelligent practice as "an application of knowledge to instrumental decisions" (p. 50) is inadequate. When theories are conflicting or incomplete, developing a theory of professional practice, according to Argyris and Schon (1974), requires problem-setting or diagnosis, testing, and belief in personal causation (p. 158).

Diagnosis. Diagnosis or problem setting is the ability to set or frame the problem based on professional knowledge, past experience, the uniqueness of the situation and people involved, social and professional norms of behavior, and expectations held by others. Technical rationality, on the other hand, emphasizes problem solving by concentrating on clearly defined ends achieved through means dictated by professional knowledge.

The practitioner draws upon past experiences and understandings in setting the unique problem. Problem setting is the process in which "we name the things to which we will attend and frame the context in which we will attend to them" (Schon, p. 40). Past experience is useful in framing problematic situations because the new situation can be seen as similar to past situations or combinations of past situations, or as different than past situations. Problem framing or setting is necessary to create a manageable situation of a problematic one. An example of a Diagnosis item is "I find many of my own early school experiences useful in managing my students."

Testing. Once a problem is framed, the reflective practitioner takes an exploratory stance. The reframing of the problem is evaluated according to its ability to be solved, the desirability of the solution, and the congruence of the solution with fundamental values and theories held by

the practitioner.

Problem setting and testing require that the practitioner engage in a sense-making analysis of an ambiguous situation. In this regard, it is necessary that the professional be comfortable approaching the situation as problematic. Budner (1962) defines an ambiguous situation as one characterized by "novelty, complexity, or insolubility" (p. 30). An individual who is threatened by ambiguity will react with submission or denial; i.e., he or she will accept the situation as insoluble or will perceive it unrealistically to fit expectations. Budner developed a sixteen-item scale to measure an individual's tolerance of ambiguity. Since problem-setting and testing require an openness to novelty and complexity, it is expected that reflective practitioners will report high tolerance of ambiguity.

Personal Causation. The third requirement for developing an effective theory of practice is personal causation. The practitioner must be committed to the personal and professional values used in setting the problem and accept responsibility for actions taken. Unless there is a strong commitment to values and to self, the practitioner will be unable to question the conventions of the profession where necessary, will have difficulty admitting perceived failure, and will resist testing.

Personal causation is the acceptance of responsibility for actions and their consequences. Argyris and Schon (1974) find it to be implicit in the willingness to test assumptions openly. Unless the practitioner is committed to the values and theories from which the problem is framed and on which action is based, innovative solutions are unlikely. Thus, a reflective practitioner would respond favorably to items such as "If students are having trouble in school, it's up to the teacher to find the solution."

Bandura's (1982) self-efficacy theory is similar to personal causation in that self-efficacy is the belief that one has the requisite skills necessary to bring about desirable outcomes in a given situation. Whereas other instruments have been developed to measure teacher control expectancies concerning student outcomes (Maes & Anderson, 1985), the Teacher Efficacy Scale (Gibson & Dembo, 1984; Dembo & Gibson, 1985) was designed to measure 1) the extent to which teachers feel that they, as opposed to external factors, control student outcomes, and 2) the amount of confidence the teacher has in his or her own abilities to successfully influence student achievement. The scale is thus useful in validation of the construct of personal

causation.

METHOD

The Reflective Teaching Instrument was developed around these three dimensions of reflective practice - diagnosis, testing, and personal causation. Instrument development consisted of four phases: 1) item generation based on indicators derived from a review of related literature, 2) assessment of face validity and item revision by an expert panel, 3) pilot testing to determine content and construct validity, and 4) a field study of the revised scale to assess reliability and construct validity. This instrument development procedure is based on selected methods used by Okeafor (1983) in the development of a logic of confidence scale, and is similar to the Likert technique for affective scale development described by Anderson (1981).

Indicators of each of the three dimensions of reflective practice were compiled from a review of the literature on reflective teaching and experiential learning. One or more items believed to tap each indicator were generated, resulting in a total of 60 items with twenty items each representing diagnosis, testing, and personal causation.

Three professors of educational administration and research and one doctoral student in education familiar with the conceptual framework of the study comprised an expert panel that assessed the face validity of each item. Using a Q-sort technique, each panel member grouped the items according to the concept they were believed to tap. Items that panelists believed to overlap two or more categories were sorted to the category deemed best. All items sorted to the same concept by at least three of the four experts were retained for the pilot test. Twelve items with less than 75% agreement were eliminated during this phase.

The pilot instrument thus consisted of 48 six-point Likert-format items. The 16 items from Budner's (1962) Tolerance/Intolerance of Ambiguity Scale and the 16 items from Gibson and Dembo's (1984) Teacher Efficacy Scale were appended to the Reflective Teaching pilot instrument for preliminary construct validation purposes.

RESULTS

To assess the internal reliability and construct validity of the pilot instrument, four professors of education were solicited to administer the instrument to practicing elementary and secondary school teachers enrolled

in their graduate classes during the spring semester of 1987. Forty of these practicing teachers voluntarily completed the pilot instrument. To avoid sensitization to the constructs under consideration, the instrument received by participants did not include identification of sub-scales and was entitled only "RTI".

A teacher's reflective practice score was determined by summing the scores on diagnosis, testing, and personal causation items. Items were selected for inclusion in the field study based on analysis of pilot results. Two criteria were used for item retention: large (standard deviation greater than 1.0) variability in item response (Nunnally, 1978, pp. 281-282) and the Likert criterion of internal consistency of affective scales (Anderson, 1981, pp. 248-249).

Biserial correlations of items with total test score is an appropriate indicator of content validity and reliability, according to Anderson (1981). Twenty-three items with significant positive correlation to total score ($r \geq .30$, $p < .05$) were retained. Three additional items with standard deviations greater than 1.0 were selected on the basis of their acceptable variability in response and significant correlations ($p < .05$) to their respective sub-scale scores. Alpha reliability for the pilot sample (N=40) on the 26 items of the revised scale was estimated at .78.

In addition to assessment of face validity by the expert panel, construct validity of two components of reflective practice was supported in the pilot study. As expected, the sum of the ratings of the nine items measuring diagnosis was significantly correlated ($r = .47$, $p < .01$) with tolerance of ambiguity as was the sum of the 11 items measuring testing ($r = .31$, $p < .05$). The sum of item ratings on the personal causation construct was significantly correlated ($r = .52$, $p < .001$) with teacher efficacy.

Having reduced the RTI to 26 items, a field study was conducted with a larger teacher sample. The instrument was further refined based on field study results.

The field study sample consisted of 108 public school elementary and junior high teachers of reading, math, and English. Fifty-four teachers were randomly chosen from nine elementary schools, and 54 from six junior high schools. All schools were located in a large suburban school district. One hundred two teachers returned usable instruments, a response rate of 94%.

Reliability. Alpha reliability coefficients of the Reflective Teaching Instrument in the pilot study of 40 subjects was .78. Reliability based on the 102 participants in the field study was .65. Although reliability coefficients must be interpreted relative to the length of the test and the degree of measurement error that is deemed acceptable, Nunnally (1978) advises that a coefficient of used to measure hypothesized constructs.

Although it is expected that reliability of the sub-scales would be lower than that of the total scale due to the smaller number of items, the low reliability (alpha=.31) of the diagnosis sub-scale indicated some problems with items on this scale, confirmed through factor analysis.

Validity. To determine the construct validity of the Reflective Teaching Instrument, factor analytic techniques were applied to the items of the total scale. According to Anastasi (1966) and Crocker and Algina (1986), factor analysis is appropriate in the construct validation of a test believed to measure a theoretical construct composed of a group of underlying constructs. A three-factor solution was obtained through principal components extraction with Varimax rotation.

Two decision rules were used for item retention. All items had to have factor loadings of at least .35 on the factors they were believed to represent, and no item could have a higher loading on a different factor. This yielded a 15-item scale that is theoretically sound and empirically demonstrated by a second factor analysis.

Again using principal components extraction with Varimax rotation, factor structure coefficients were obtained for the revised 15-item scale (see Table 1). Again, all items had to meet the criterion that their factor loadings be approximately .35 or greater on their intended factors. A scree plot (Cattell, 1952) of the eigenvalues of all factors supported the acceptance of a three-factor model. The three factors representing Testing, Diagnosis, and Personal Causation, respectively, account for 44.6% of the variance in item response.

Insert Table 1 about here

The items empirically supported through factor analysis to constitute the three dimensions of reflective teaching

were retained. Four of the original Diagnosis items, five of the Testing items, and all six Personal Causation items comprise the revised 15-item scale.

Face validity of the Reflective Teaching Instrument was assessed in the pilot study. Content validity was inferred from the high reliability of the instrument in the pilot study (alpha=.78). However, the alpha reliability in the field study dropped to .65 with two of the three sub-scale alphas below .50. On the basis of factor analysis, the scale was reduced to 15 items. Alpha coefficients for the revised scale using the field sample are presented in Table 2.

Insert Table 2 about here

The increased reliability coefficients of the revised scale and its sub-scales support the content validity of the RTI; i.e., the items appear to measure the identified constructs within each sub-scale.

The concurrent validity of two sub-scales of the RTI was assessed in the pilot study by correlation with two other scales believed to tap similar constructs. Tolerance of ambiguity (Budner, 1962) was positively correlated with both Diagnosis and Testing in the 40-subject pilot study. Likewise, it was found that teacher efficacy (Gibson & Dembo, 1984) and Personal Causation on the RTI were positively correlated.

The Pearson correlation in the 102-subject field study between tolerance of ambiguity and Diagnosis on the RTI was and the Testing sub-scale remained significant at .23 ($p<.05$). The correlation between teacher efficacy and Personal Causation in the field study was .55 ($p<.0001$).

CONCLUSIONS

All research involving the construct of reflective practice has heretofore consisted of case studies or qualitative data collection and analysis, primarily due to the inability to describe reflective practice. Like Sergiovanni's (1982) theory of leadership, reflective practice has not been tested in causal comparative research because it has not been or can not be operationalized.

The applicability of the theory of reflective practice

depends upon its ability to be empirically tested in a series of studies, studies requiring operational definitions of key constructs. A theory is of little value if the major constructs are too obscure to allow development of any satisfactory operational definition. Argyris and Schon's (1974) theory has been particularly suspect to this criticism.

Educators striving to gain professional status must clarify the theories they espouse less they continue to be criticized for disguising critical issues in a cloak of jargon. Our theories must stand up to empirical scrutiny. A preliminary instrument that can be used in empirical tests of the theory of reflective practice does now exist.

One question that emerges with regard to the findings of this research concerns the theory of reflective practice itself. The heuristic value of the theory, that is, its ability to stimulate research (Kerlinger, 1979) is called into question by the vague nature of its components. Investigation of reflective teaching requires operationalizing key constructs whose exact meanings are poorly articulated.

Kerlinger (1979) defines theory as "a set of interrelated constructs (concepts), definitions and propositions that present a systematic view of phenomena by specifying relationships among variables, with the purpose of explaining and predicting the phenomena" (1986, p. 9). This definition requires that the constructs be defined, yet Schon's (1983) definition of reflective practice is presented in terms of other constructs also requiring definition.

Future research should include further refinement of the instrument with particular attention to the Diagnosis sub-scale whose construct validity is somewhat suspect due to its low correlation with tolerance of ambiguity. Research is also needed that investigates the relationship between reflective practice and teacher effectiveness. A positive relationship between scores on the RTI and measures of teacher effectiveness would provide much-needed support for teacher education programs that are inquiry oriented, and would greatly increase our understanding of the research begun by Berliner (1986) into differences between the expert and novice pedagogue.

Reflective practice assumes a need to select from all presenting problems that particular frame to which one will attend (i.e., diagnosis). For this reason, the relationship between reflective practice and effectiveness might be

examined among educators whose jobs appear less technically rational, or more ambiguous. Principals, typically held accountable for any number of non-specific tasks, would represent a unique population to examine the diagnosis and testing dimensions of reflective practice.

Finally, reflective teaching in different school and classroom contexts warrants consideration. If a relationship between reflective teaching and teacher effectiveness is supported, reflective teaching should be encouraged and taught to interns. It must first be clear, however, that this relationship applies in all school contexts. The effect of reflective teaching on teacher effectiveness should be examined at different subject and grade levels, and particularly, in classrooms posing unique instructional problems.

References

Anastasi, A. (1976). Pychological testing. New York: Macmillan Publishing Co.

Anderson, L. W. (1981). Assessing affective characteristics in the schools. Boston: Allyn and Bacon, Inc.

Argyris, C. & Schon, D. A. (1974). Theory in practice: Increasing professional effectiveness. San Francisco: Jossey-Bass, Inc.

Bandura, A. (1982). Self-efficacy machanism in human agency. American Psychologist, 37(2), 122-147.

Berliner, D. C. (1986). In pursuit of the expert pedagoque. Educational Researcher, 15(7), 5-13.

Budner, S. (1962). Intolerance of ambiquity as a personality variable. Journal of Personality, 30, 29-50.

Cattell, R. B. (1952). Factor analysis. New York: Harper and Row.

Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. New Tork: Holt, Rinehart and Winston.

Dembo, M. H. & Gibson, S. (1985). Teachers' sense of efficacy: An important factor in school improvement. The Elementary School Journal, 2, 173-184.

Gibson, S. & Dembo, M. H. (1984). Teacher efficacy: A construct validation. Journal of Educational Psychology, 4, 569-582.

Kerlinger, F. N. (1979). Behavioral research: A conceptual approach. New York: Holt, Rinehart, and Winston.

Korthagen, F. A. (1985). Reflective teaching and preservice teacher education in the Netherlands. Journal of Teacher Education, 36, 11-15.

MacKinnon, A. M. (1986, April). Detecting reflection-in-action in preservice elementary science teachers. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Maes, W. R., & Anderson, D. E. (1985). A measure of teacher locus of control. Journal of Educational Research, 79(1), 27-32.

Nunnally, J. (1978). Psychometric theory. New York: McGraw-Hill.

Oberg, A. & Field, R. (1986, April). Teacher development through reflection on practice. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Okeafor, K. R. (1983). The logic of confidence: Toward definition and measurement. (Doctoral dissertation, Ohio State University, 1983). Dissertation Abstracts International, 44, 3363-A.

Russell, T. L. (1985). Images of improving practice. Teacher Education Quarterly, 12(3), 16-22.

Russell, T. L. (1986, April). Beginning teachers' development of knowledge-in-action. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Schon, D. A. (1983). The reflective practitioner: How professionals think in action. New York: Basic Books.

Schon, D. A. (1987, April). Educating the reflective practitioner. Invited address to the meeting of the American Educational Research Association, Washington, D.C.

Table 1

Factor Structure of the Reflective Teaching Instrument

Revised Scale - 15 items

Item	Factor 1	Factor 2	Factor 3
Diagnosis 1	0.1057	0.3498	0.0321
Diagnosis 2	-0.0036	0.5422	0.0104
Diagnosis 4	0.0426	0.7515	0.1384
Diagnosis 5	0.0576	0.8047	0.0707
Testing 3	0.7333	0.0125	-0.0206
Testing 5	0.6014	0.2727	0.0780
Testing 8	0.5819	0.4364	0.0231
Testing 10	0.6644	-0.3393	0.0764
Testing 11	0.6834	0.3803	0.2715
Per.Caus. 1	0.2659	0.1968	0.4809
Per.Caus. 2	0.0799	0.2224	0.6358
Per.Caus. 3	-0.0007	0.1809	0.6737
Per.Caus. 4	0.3297	-0.1179	0.3792
Per.Caus. 5	0.0357	0.0014	0.6640
Per.Caus. 6	-0.0186	-0.0930	0.5072

Table 2

Alpha Reliability of the Reflective Teaching Instrument 15-Item Scale

Sub-scale	i	Alpha
Diagnosis	4	.47
Testing	5	.69
Personal causation	6	.60
Total scale	15	.70

Note. i = # of items
n = 102